The OpenAI Copyright Lawsuit: Could It Backfire on Canadian Media?

BY ISMINE OSMAN / ON MAY 12, 2025 / IN INTELLECTUAL PROPERTY, UNCATEGORIZED

By: Ismine Osman

Matter Commented On: *Canadian News Media Companies v OpenAI*, Statement of Claim

PDF Version: The OpenAI Copyright Lawsuit: Could it Backfire on Canadian Media?

Introduction: A Legal Paradox for Canadian Media

In November 2024, a group of Canada's largest news media companies (Plaintiffs), including the Toronto Star,

Metroland Media Group, Postmedia, The Globe and Mail, The Canadian Press, and CBC/Radio-Canada, filed a lawsuit against OpenAI (*Canadian News Media Companies v OpenAI*, Statement of Claim (28 November 2024) (Statement of Claim)). They allege that OpenAI scraped and copied content without consent to train its artificial intelligence (AI) models (Statement of Claim at paras 44-45). The Plaintiffs also claim that OpenAI's models may reproduce parts of this content in user-facing outputs, which could further support the allegation of infringement (Statement of Claim at para 5). Legal commentators, including Michael Geist and Howard Knopf, have already weighed in on the lawsuit's weaknesses and strategic undertones (see Howard Knopf, "AI Litigation for the Canadian Nation"; Michael Geist, "Canadian Media Companies Target OpenAI in Copyright Lawsuit But Weak Claims Suggest Settlement the Real Goal").

This blog explores whether such training amounts to unauthorized reproduction under section 3(1) of the *Copyright Act*, RSC 1985, c C-42 (*Copyright Act*), and whether OpenAI's use qualifies as fair dealing under section 29. Ironically, a ruling in the Plaintiffs' favour could also limit how Canadian media rely on fair dealing when using third-party content in their own reporting. To assess those questions, I first examine the core legal issue at stake: reproduction.

The Legal Claims Against OpenAI: The Reproduction Question

A key legal issue in this case is whether training AI models on copyrighted news content qualifies as "reproduction" under the *Copyright Act*. Section 3(1) of the *Copyright Act* gives copyright holders the exclusive right to reproduce their works or any substantial part of them. The Plaintiffs argue that OpenAI's models copied and reproduced their content without authorization, which would violate these rights (Statement of Claim at paras 52–54).

Much will depend on how the training process is technically implemented, specifically whether expressive content is retained or simply analyzed for statistical patterns. In *CCH Canadian Ltd. v. Law Society of Upper Canada*, 2004 SCC 13 (CanLII) (*CCH*), the Supreme Court of Canada (SCC) clarified that copyright protects the expression of ideas, not the ideas or facts themselves and that original expression must reflect skill and judgment (*CCH*, at paras 14,16). The SCC also noted that copying, even in full, may be allowed if it meets the test for fair dealing (*CCH,* at para 56).

In the context of determining whether AI training constitutes reproduction under the *Copyright Act*, OpenAI's interaction with copyrighted news content could take two different forms, each raising distinct legal questions.

1. Direct Reproduction

If OpenAI copied news articles or substantial parts of them word-for-word, this would qualify as reproduction

under section 3(1) of the *Copyright Act*. In *CCH*, the SCC confirmed that copyright holders have the exclusive right to reproduce their works or any substantial part thereof under section 3(1), and that doing so without consent may constitute infringement under section 27(1) unless a statutory exception such as fair dealing applies (*CCH*, at paras 11–12). Even partial copying can infringe if it captures an embodied part of a work that reflects the author's skill and judgment (*CCH*, at paras 14, 16).

2. Extracting Patterns Without Retaining Content

OpenAI may argue that its models do not store or reproduce copyrighted works. Instead, they analyze patterns in large datasets, such as word frequency or sentence structure, without retaining expressive content fixed in a work. Since copyright protects only the expression of ideas embodied in a work and not the ideas themselves, such pattern extraction may fall outside the scope of reproduction (*CCH*, at para 8).

The Plaintiffs dispute this view. The Statement of Claim alleges that OpenAI copied datasets in full or in part without authorization and that such copying involved expressive content, not just abstract patterns (Statement of Claim at paras 44–45). Professor Michael Geist explains that one of the contested legal questions will be whether tokenization, which involves converting text into numerical representations for statistical analysis, constitutes copying or merely analyzing underlying works without reproducing them (Geist, 2024).

This issue has also caught the attention of policymakers. In its 2025 report (see here), the Government of Canada acknowledged the need for more fact-finding to understand if and how copies are made during AI training (Government of Canada, at 6). This reinforces that the legal and technical boundaries of reproduction in the AI context remain uncertain and continue to evolve.

Whether courts ultimately find that OpenAI's practices amount to reproduction will likely depend on technical evidence, specifically whether the models retain expressive content in a permanent or fixed form or generate outputs that reproduce substantial parts of the original news sources. But even if reproduction is found, the question remains whether OpenAI can rely on fair dealing as a defence.

Fair Dealing and AI: Can OpenAI Use the Same Defence as Canadian Media?

If training AI models on news content is found to involve reproduction, the next question is whether OpenAI's actions qualify as fair dealing under section 29 of the *Copyright Act*. Fair dealing is a user's right that permits the use of copyrighted material without permission in certain contexts, like research or education. The Statement of Claim anticipates that OpenAI may attempt to rely on fair dealing but argues that any such defence should fail because the use was not fair, not for an allowable purpose, and not carried out by a user (Statement of Claim at para 54). This raises an interesting tension. While Canadian media companies

rely on fair dealing to summarize and republish third-party content for news reporting, they argue that similar uses by AI developers fall outside the law.

Even if courts conclude that OpenAI reproduced copyrighted content during model training, the company could argue that the training process qualifies as "research," one of the purposes listed under section 29. In *CCH*, the SCC confirmed that fair dealing is a user's right and must be interpreted broadly to promote access to knowledge by enabling users to engage with copyrighted works for legitimate purposes such as research, private study, and education (at para 48). This principle could support the view that using copyrighted materials to train AI models ultimately serves the public interest by improving access to knowledge through more effective information tools.

The Court also held that "research" should receive a large and liberal interpretation and is not limited to personal or non-commercial uses (at para 51). For example, lawyers working for profit were still found to be engaged in research when advising clients or preparing legal arguments (at para 51).

To assess whether a dealing is fair, courts apply a six-factor test set out in *CCH* (at para 53). These factors include the purpose, character, amount, alternatives, nature, and effect of the dealing. The following section considers each factor in turn, focusing on how they apply to OpenAI's model training.

1. Purpose of the Dealing

The first question is whether the purpose qualifies under the *Copyright Act*. The Court in *CCH* said that "research" includes commercial contexts, not just personal study (at para 51). OpenAI may argue that training language models qualifies as research because it involves analyzing language patterns to build useful tools rather than selling or distributing the original content. This position mirrors the reasoning in *SOCAN v. Bell Canada*, 2012 SCC 36 (CanLII) where the SCC found that consumer music previews qualified as fair research under a broad interpretation of section 29 (at para 27). OpenAI's use involves extracting text from large volumes of online content to train predictive models, which it may frame as necessary for improving information tools. Still, it remains unsettled whether courts will accept AI training as research in this context.

2. The Character of the Dealing

This factor looks at how the material was copied and how widely it was distributed. Copying that is limited in scope, used for a short time, or confined to internal use without public disclosure tends to support fairness. Systematic or large-scale reproduction makes it harder to justify. If OpenAI can show that it used text to train language models without retaining expressive content in a fixed form or making the original articles accessible to users, this may support its position. However, copying large volumes of news content could weigh against a finding of fairness (*CCH*, at para 55).

3. How Much Was Copied

Courts consider both how much of the work was copied and whether the portion used was central or essential to the original. Even short excerpts can be problematic if they include the "heart" of the work (*CCH*, at para 56). If OpenAI ingested full articles or large sections, that could be difficult to defend. On the other hand, if it only used brief or random snippets to detect patterns and avoided preserving original expression, this may favour a finding of fairness. The Statement of Claim also emphasizes that OpenAI has monetized its GPT models through paid offerings such as ChatGPT Plus and Team, earning significant revenue (Statement of Claim at paras 48–49).

4. Whether There Were Alternatives

Courts also ask whether the same outcome could have been achieved without copying the material (*CCH*, at para 57). OpenAI may argue that training effective language models requires exposure to large volumes of real-world text and that there is no viable alternative. Plaintiffs might argue that OpenAI could have used licensed data or public domain content instead.

5. Nature of the Work

This factor looks at whether the work was private or meant to be widely shared. Publicly accessible material, like news articles, is more likely to support fair dealing (*CCH*, at para 58). Since OpenAI reportedly used mainly public sources, this could work in its favour. However, if

some of the data came from subscription-based or paywalled platforms, courts may see that as tipping the balance the other way.

6. Whether the Use Hurts the Market

This final factor considers whether OpenAI's use harms the market for the original work. If AI-generated responses act as substitutes for news reporting or reduce web traffic and advertising revenue, this may strongly weigh against fairness. While Canadian fair dealing and U.S. fair use are distinct legal doctrines, both consider the impact on the market for the original. In *Thomson Reuters Enterprise Centre Gmbh v Ross Intelligence Inc.*, the U.S. District Court for the District of Delaware found that training an AI model on Westlaw's copyrighted legal headnotes harmed the market for legal research tools and weighed against fair use (No. 1:20-cv-613-SB, D. Del.., Feb. 11, 2025). Although not binding in Canada, the case illustrates how courts may approach market harm in the context of AI model training.

Canadian courts, however, place particular emphasis on the evidentiary record when assessing this factor. In *CCH*, the SCC noted that the copyright holders failed to provide evidence that the alleged copying negatively affected the market for their legal publications (at para 72). In fact, the Court emphasized that the plaintiffs continued to publish and sell new works during the period of alleged infringement. This highlights a crucial point, which is that while the defendant bears the burden

of proving fair dealing, courts will expect the copyright owners to produce market harm evidence if it exists.

This emphasis on market effects also underscores the evidentiary burden that exists in Canadian law. As the SCC confirmed in *York University v. Access Copyright*, 2021 SCC 32 (CanLII) fair dealing must be more than just claimed; it must be proven with evidence (at paras 96–100). For OpenAI, this means offering a clear explanation of how its training process limits copying, avoids replacing original content, and aligns with the goals of the *Copyright Act*.

Broader Implications and Closing Thoughts

The OpenAI lawsuit could reshape how Canadian copyright law applies to AI. If OpenAI succeeds, courts may adopt a broader interpretation of "research" under fair dealing, giving developers more flexibility to train models using copyrighted content. That outcome could support innovation and align with how courts have historically approached evolving technologies. However, it may also weaken the ability of media companies and other sectors that rely on exclusive control over their content to regulate how their materials are used in digital environments.

If OpenAI loses, courts could require companies to obtain licences before using copyrighted material for AI training. This may slow technological development and increase costs, not only for developers but also for industries that depend on text and data to build tools,

conduct research, or deliver services. At the same time, stricter rules could affect how Canadian media organizations use third-party content in their own reporting. If courts narrow the scope of fair dealing in response to AI, journalists and other creators may face new uncertainty when referencing or reproducing the work of others.

There is also a chance that a ruling against OpenAI could lead the federal government to step in and introduce a new exception to copyright infringement for AI model training, especially if the court signals that the current law does not address this new technology properly.

This goes back to the SCC's emphasis in *CCH* that fair dealing must remain flexible enough to respond to new technologies (at para 60). This case now tests whether that flexibility can extend to the training of generative AI systems. Regardless of the outcome, the decision is likely to influence how developers, creators, and policymakers understand the role of copyright in the age of AI.

This post may be cited as: Ismine Osman, "The OpenAI Copyright Lawsuit: Could It Backfire on Canadian Media?" (12 May 2025), online: ABlawg, http://ablawg.ca/wp-content/uploads/2025/05/Blog_IO_OpenAICopyright.pdf

To subscribe to ABlawg by email or RSS feed, please go to http://ablawg.ca

Follow us on Twitter @ABlawg

PREVIOUS

An Important Alberta Crown Lease Continuation Decision

NEXT

The Information Commissioner's Report on the Government of Alberta's War on the Public's Right to Access Information

POWERED BY WORDPRESS & THEME BY ANDERS NORÉN